Overview – Chemical and Scoring Ranking Assessment Model

SCRAM: A Scoring and Ranking System for Persistent, Bioaccumulative, and Toxic Substances for the North American Great Lakes *

Part I: Structure of the Scoring and Ranking System *(ESPR No. 1, 2000)* **DOI: http://dx.doi.org/10.1065/espr199910.009**
Part II: Bioaccumulation Potential and Persistence *(ESPR No. 2, 2000)* **DOI: http://dx.doi.org/10.1065/espr199910.010**
Part III: Acute and Subchronic or Chronic Toxicity *(ESPR No. 3, 2000)* **DOI: http://dx.doi.org/10.1065/espr199910.011**
Part IV: Results from Representative Chemicals, Sensitivity Analysis, and Discriminatory Power *(ESPR No. 4, 2000)* **DOI: http://dx.doi.org/10.1065/espr199910.012**

Erin M. Snyder [1], Shane A. Snyder [1], John P. Giesy [1]
Shari A. Blonde [2], Gary K. Hurlburt [2,3], Cheryl L. Summer [2], Rachel R. Mitchell [2], Dennis M. Bush [2]

[1] Department of Zoology, Michigan State University, National Food Safety and Toxicology Center and Institute of Environmental Toxicology, East Lansing, MI 48824, USA
[2] Surface Water Quality Division, Michigan Dept. of Environmental Quality, PO Box 30273, Lansing, MI 48909-7773, USA
[3] Deceased

Corresponding author: Erin M. Snyder

Part I: Structure of the Scoring and Ranking System

Hundreds of chemical contaminants have been identified in the Great Lakes System of North America. Depending on the agency or organization, various subset lists of these contaminants have been identified as chemicals of potential concern. However, there is no agreement on the method that should be used to make management decisions. Except for consensus on approximately 40 chemicals that most North American agencies agree can cause deleterious effects if released into the environment, no agreement has been reached regarding the priority that contaminants should receive for further action. That leaves hundreds of chemicals that have been, are being, or potentially could be released into the environment that have not been evaluated yet. A profile for potential chemicals of concern is generally thought to include persistence in the environment, potential to bioaccumulate, and ability to cause toxic effects at environmentally relevant concentrations. Except for the International Joint Commission's definition of persistence (> 8 weeks residence time in air, water, soil or sediment), there is little concurrence about what defines these characteristics. For instance, the State of Michigan currently has no established definitions or profiles of persistent, bioaccumulative, toxic substances. Furthermore, there is no standard process to rank chemicals relative to these characteristics. The Chemical Scoring and Ranking Assessment Model (SCRAM) has been developed to provide a process to rank-order chemicals based on these characteristics. The SCRAM system was developed primarily for use in the Great Lakes region of North America and particularly in Michigan, but it is not site-specific. Use of this system may assist in pollution prevention activities and other future chemical control efforts, allowing attention to be focused first on those chemicals likely to present the greatest hazard.

Part II: Bioaccumulation Potential and Persistence

Part I of this series introduced SCRAM, a chemical scoring and ranking system for contaminants of the North American Great Lakes. Here, in **Part II**, scoring of the bioaccumulation potential and persistence of chemicals is discussed, including acceptable types of data, specific scoring instructions, and the basis for criteria and scores for these categories of the system. Difficulties encountered during the process of determining which types of data adequately represent the properties of interest are discussed. Also, justification is given for an emphasis on scoring on the basis of persistence.

Part III: Acute and Subchronic or Chronic Toxicity

In **Part II**, scoring of the potential for a chemical to persist in the environment and bioaccumulate was described. In **Part III**, scoring of chemical toxicity is discussed, including definitions and descriptions of effects that are scored, specific scoring instructions, the basis for the criteria and scores, and specific conditions or concerns regarding the types of data used for scoring. A score for each chemical screened is determined from available test data from acute or subchronic and chronic toxicity tests conducted on aquatic and terrestrial organisms. Subchronic and chronic human health effects, including carcinogenicity, are also considered. Part IV includes an evaluation of the performance of the scoring and ranking system.

Part IV: Results from Representative Chemicals, Sensitivity Analysis, and Discriminatory Power

The Chemical Scoring and Ranking Assessment Model (SCRAM) has been described in **Parts I-III** of this series. SCRAM is a chemical scoring and ranking (CSR) system that scores chemicals on the basis of bioaccumulation potential, environmental persistence, and toxicity. **Part IV** describes various tests and descriptions of the performance of this system. A group of 21 representative chemicals was chosen and scored to test the system. For those chemicals, the percentages of the scores associated with fate-related properties and associated with data uncertainty were determined. The scoring of four of these chemicals is described in greater detail, and the suitability of the scores is discussed. An analysis of the sensitivity of the system to incomplete data sets is presented. And finally, the discriminatory power of the system is described.

* The scoring and ranking system in the form of a Lotus 123[97] spreadsheet and a description of its use are available on the Internet at http://www.epa.gov/toxteam/pbtrept/

SCRAM: Chemical Scoring and Ranking Assessment Model

SCRAM: A Scoring and Ranking System for Persistent, Bioaccumulative, and Toxic Substances for the North American Great Lakes *

Part I: Structure of the Scoring and Ranking System
Part II: Bioaccumulation Potential and Persistence
Part III: Acute and Subchronic or Chronic Toxicity
Part IV: Results from Representative Chemicals, Sensitivity Analysis, and Discriminatory Power

Part III. Acute and Subchronic or Chronic Toxicity

Erin M. Snyder[1], Shane A. Snyder[1], John P. Giesy[1]
Shari A. Blonde[2], Gary K. Hurlburt[2,3], Cheryl L. Summer[2], Rachel R. Mitchell[2], Dennis M. Bush[2]

[1] Dept. of Zoology, Michigan State University, National Food Safety and Toxicology Center and Institute of Environmental Toxicology, East Lansing, MI 48824, USA
[2] Surface Water Quality Division, Michigan Dept. of Environmental Quality, PO Box 30273, Lansing, MI 48909-7773, USA
[3] Deceased

Corresponding author: Erin M. Snyder

DOI: http://dx.doi.org/10.1065/espr199910.011

Abstract. In Part I of this series (Snyder et al., 1999a), the Chemical Scoring and Ranking Assessment Model (SCRAM) was introduced. This system produces scores for chemicals based on their bioaccumulation potential, environmental persistence, and toxicity. In Part II, scoring of the potential for a chemical to persist in the environment and bioaccumulate was described (Snyder et al., 1999b). In Part III, scoring of chemical toxicity is discussed, including definitions and descriptions of effects that are scored, specific scoring instructions, the basis for the criteria and scores, and specific conditions or concerns regarding the types of data used for scoring. A score for each chemical screened is determined from available test data from acute or subchronic and chronic toxicity tests conducted on aquatic and terrestrial organisms. Subchronic and chronic human health effects, including carcinogenicity, are also considered. Part IV includes an evaluation of the performance of the scoring and ranking system (Snyder et al., 1999c).

Keywords: Acute toxicity; bioaccumulation; chemical scoring and ranking; chronic toxicity; hazard; North American Great Lakes; persistence; priority pollutants; SCRAM (Chemical Scoring and Ranking Assessment Model); uncertainty; water pollution

1 Introduction

In **Part I** of this series, the Chemical Scoring and Ranking Assessment Model (SCRAM) was introduced (Snyder et al., 1999a). This system produces scores for chemicals based on their bioaccumulation potential, environmental persistence, and toxicity. In **Part II**, scoring of the potential for a chemical to persist in the environment and bioaccumulate was described (Snyder et al., 1999b). In **Part III**, scoring of chemical toxicity is discussed. A score for each chemical screened is determined from available test data from acute and chronic tests conducted on aquatic and terrestrial organisms. Subchronic and chronic human health effects, including carcinogenicity, are also considered. If a chemical is determined by SCRAM to have relatively low environmental persistence, it is scored for toxicity on the basis of acute toxicity test data. If the chemical is relatively persistent, it is scored for toxicity on the basis of subchronic or chronic toxicity test data. A literature search was conducted for a list of 21 test chemicals to determine whether the types of data required for scoring were available and to ensure that the scoring ranges and criteria were appropriate. The chemicals were scored to test the system, and the scoring is described in **Part IV** of this series (Snyder et al., 1999c). Data selection criteria for toxicity scoring are described here. **Table 1** (→ *p. 3*) gives abbreviations and definitions of terms used.

2 Data Selection Criteria for Acute and Subchronic/Chronic Toxicity

Studies were considered to be acceptable if they followed published testing guidelines or were conducted in Good Laboratory Practices (GLP)-certified tests using standardized guidelines. Examples of guidelines for aquatic studies are the American Society for Testing and Materials (ASTM) guidelines: the

* The scoring and ranking system in the form of a Lotus 123[97] spreadsheet and a description of its use are available on the Internet at http://www.epa.gov/toxteam/pbtrept/

Table 1: Abbreviations and definitions of terms

EC50	effect concentration 50%, or concentration of a chemical that is required to elicit a particular effect in 50% of a test population within a specified time period (RAND, 1995)
ED50	effect dose 50%, or dose that produces a particular effect in 50% of the test population within a specified period of time (RAND, 1995)
LC50	lethal concentration 50%, or concentration of a chemical in water to which test organisms are exposed that is estimated to be lethal to 50% of the test population within a specified period of time (RAND, 1995).
LD50	lethal dose 50%, or dose that is lethal to 50% of the test population within a specified period of time
LOAEC	lowest observed adverse effect concentration
LOAEL	lowest observed adverse effect level
LOEC	lowest observed effect concentration
LOEL	lowest observed effect level
MATC	maximum acceptable toxicant concentration, usually the geometric mean of the LO(A)EC and the NO(A)EC
NOAEC	no observed adverse effect concentration
NOAEL	no observed adverse effect level
NOEC	no observed effect concentration
NOEL	no observed effect level

ASTM Standard Guide for Conducting Acute Toxicity Tests with Fishes, Macroinvertebrates and Amphibians (ASTM, 1992a); the ASTM Standard Guide for Conducting Early Life-Stage Toxicity Tests with Fishes (ASTM, 1992b); or the ASTM Standard Guide for Conducting Renewal Life-Cycle Toxicity Tests with *Daphnia magna* (ASTM, 1992c). For the terrestrial toxicity and human health portions of the system, the Organization for Economic Cooperation and Development (OECD) Guidelines for Testing Chemicals (OECD, 1981) and U.S. Environmental Protection Agency (U.S. EPA) Toxic Substances Control Act (TSCA) Test Guidelines (U.S. EPA, 1985) are examples of suitable testing guidelines.

3 Acute Toxicity

3.1 Definition and discussion of effects; specific scoring instructions

The SCRAM scoring system evaluates acute toxicity of chemicals to aquatic and terrestrial organisms. Data on acute toxicity represent the largest body of toxicity data available and provide information on the relative hazards of chemicals in the event of an accidental large-scale release. EC50 and LC50 data are used for scoring acute aquatic toxicity, and ED50 and LD50 data are used for scoring acute terrestrial toxicity. The EC50 or ED50 endpoint for use in SCRAM is effective immobilization. Data on rats, mice, monkeys, and guinea pigs can be used for scoring in the acute terrestrial toxicity category because there is no acute human toxicity category. In the subchronic/chronic terrestrial toxicity category, these data fall under the human toxicity category because these species frequently are used as surrogates for humans in toxicity testing. For both acute toxicity categories, where more than one toxicity value is available for a single subcategory, the value that would result in the greatest score is used to score the subcategory.

3.2 Basis for criteria and scores

Criteria and metrics for this scoring system were determined after reviewing existing scoring systems. During the review for the acute aquatic toxicity category, it became apparent that many of the systems were similar, with the same number of ranking tiers and endpoints. For the five acute aquatic toxicity subcategories (plants, amphibians, cold water fish, warm water fish, and invertebrates) the metrics used are the same, ranging from >1000 mg/L to <1 mg/L. The key contributors to the selection of the ranges used in SCRAM include the Oak Ridge National Laboratory (O'BRYAN and ROSS, 1988), the Michigan Critical Materials Register (CMR) (MDNR, 1987), the Ontario Ministry of the Environment (OMOE) (1990), and the United States Environmental Protection Agency (U.S. EPA) rulemaking pursuant to the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) Section 102 (U.S. EPA, 1989b). The CMR (MDNR, 1987) and Oak Ridge National Laboratory scoring systems (O'BRYAN and ROSS, 1988) are virtually identical with regard to acute aquatic toxicity scoring, with five tiers ranging from <1 mg/L to >1000 mg/L. OMOE's scoring system overlapped Oak Ridge National Laboratory's and CMR's, but added an additional scoring tier for lesser values. This did not appear to add to the discriminatory power of the system, so the five-tiered approach was used in the SCRAM system. The CERCLA document provided critical insight on acceptable endpoints and criteria required for acute aquatic toxicity evaluation.

The acute terrestrial toxicity scoring strategy was more difficult to develop since the references showed a greater range of metrics and criteria and did not use consistent parameters. For mammals, birds, reptiles, and amphibians, a combination of the most restrictive metrics was selected from the CMR (MDNR, 1987), the OMOE (1990), and the University of Tennessee (DAVIS et al., 1994) scoring systems. Metrics range from >5000 mg/kg/d to <5 mg/kg/d. These scoring systems overlapped within the five tiers that were chosen as most discriminatory for this scoring system. The CMR (MDNR, 1987) was used as the template and adjusted slightly according to the other contributors' key points. Terrestrial plants and invertebrates were not included in most of the scoring systems that were reviewed. Because it could not easily be determined how the metrics were derived, the CMR (MDNR, 1987) format was used for these subcategories because Michigan Department of Environmental Quality Staff members were familiar with the system and its derivation. The CMR (MDNR, 1987) lists the metrics as >100 mg/kg/d to <0.1 mg/kg/d for plants, and >5000 mg/kg/d to <5 mg/kg/d for invertebrates. These metrics were adjusted to fit the available data more appropriately. The units were changed to lbs/acre or kg/ha for terrestrial plants and mg/kg soil for terrestrial invertebrates because these were the units of the data most commonly available.

3.3 Specific conditions, concerns, and biases

Some difficulties were encountered during the literature search and review process, primarily due to a lack of information and the diffuse nature of pertinent data. Obtaining information on toxicity of chemicals to reptiles and amphibians was difficult since a large quantity of the data was obtained from studies that involved exposure via intraperitoneal injection instead of oral administration and so could not be considered appropriate for this system. Furthermore, the data that do exist are often of insufficient quality. It was particularly difficult to find information for terrestrial plants, terrestrial invertebrates, and reptiles and amphibians. All of these difficulties were also encountered when searching for subchronic/chronic toxicity data for these same subcategories.

Data for acute toxicity scoring usually are available for the various subcategories, but the data do not always fit the scoring requirements. In the cases of invertebrates and terrestrial plants, the initial requirements and endpoints were adjusted to fit the available data (see previous section). Terrestrial plant data were not often available in mg/kg tissue but were frequently found as application data in kg/ha or lbs/acre. Terrestrial plant data reported as mg/kg soil are converted to lbs/acre or kg/ha, assuming that the soil column is one foot deep and that soil mass is equal to 110 lb/ft^3 (range = 65-135 lb/ft^3), using the following conversions.

acre = 43,560 ft^2
acre × 0.4 = ha
1 lb = 453.6 g = 0.454 kg

Terrestrial invertebrate data reported as application data can be converted to soil concentrations (mg/kg soil) by using the same conversions. These same units and conversions also were applied to subchronic/chronic toxicity data for terrestrial invertebrates and terrestrial plants.

For amphibians, data from oral exposure studies should be scored under the acute terrestrial toxicity subcategory. Data obtained from studies using water borne exposures (e.g., FETAX bioassay, whole body tadpole exposures) should be scored under the acute aquatic toxicity subcategory.

4 Subchronic/Chronic Toxicity- Aquatic and Terrestrial Life

4.1 Definition and discussion of effects; specific scoring instructions

SCRAM evaluates the subchronic/chronic toxicity of chemicals to aquatic and terrestrial organisms. Data on subchronic/chronic toxicity represent the broadest body of toxicity data available and provide information on the relative hazards of chemicals in long-term exposure situations

4.1.1 Aquatic life

MATC, LOEC, LOAEC, NOEC, and NOAEC data are used to score subchronic/chronic aquatic toxicity to aquatic life. Hereafter, "NO(A)EC" means NOAEC or NOEC, and "LO(A)EC" means LOAEC or LOEC. Exposure periods of 28 to 33 days are considered to be adequate for reptiles, amphibians and fish. An 8-day fathead minnow study that meets American Society for Testing and Materials (ASTM) standards may be used (ASTM, 1992a). Exposure periods of 21 days for plants and for most invertebrates (e.g., *D. magna*) are adequate, but 7-day exposures for *C. dubia* are acceptable. A MATC derived with the use of uncertainty or correction factors is not acceptable for use in SCRAM.

Some subchronic/chronic aquatic toxicity test endpoints suitable for scoring are growth, reproduction, death, deformities, and any adverse effect causing immobilization of fish or amphibians (effective immobilization). For aquatic plants, some commonly used endpoints are photosynthesis, growth, and biomass.

4.1.2 Terrestrial life

LOEL, LOAEL, NOEL, and NOAEL data are used for scoring subchronic/chronic toxicity to terrestrial life. Hereafter, "LO(A)EL" means LOAEL or LOEL, and "NO(A)EL" means NOAEL or NOEL. Acceptable duration of exposure for studies on terrestrial plants (21 days for most species), invertebrates (21 days for earthworms), and amphibians (28-33 days) are the same as for aquatic species. Reptile studies of 28-33 days duration are acceptable. Data from studies of 90 days or longer duration are acceptable for mammals, and data from studies of 70 days or longer duration are acceptable for birds.

Typical endpoints used to assess subchronic/chronic terrestrial toxicity follow; however, these are not the only endpoints useful for scoring. Common endpoints available for the invertebrate subcategory include death, lethal body burden, effective immobilization, and reproduction. Terrestrial plant endpoints include photosynthesis, seed germination, growth, fresh mass, dry mass, root growth and development, survival, and production characteristics such as yield and plant height. Endpoints for mammals, birds, reptiles, and amphibians include death, deformities, reproduction, development, significant effects on growth (i.e., greater than 10% decrease in body weight) and significant effects on factors affecting viability (i.e., severe histopathological effects or severe clinical signs). Effective immobilization also is an endpoint for reptiles and amphibians. The endpoints used for assessment of toxicity in birds also include egg development and severe edema in developing chicks. Egg injection studies may be used for avian toxicity assessment. Endpoints used must be linked to population-level effects. Changes in enzyme function or hematology are not adequate endpoints for scoring. When data from whole animal studies are available, systemic, immunologic, and hormonal effects such as endocrine disruption also are examined.

The following specific scoring instructions apply to both the subchronic/chronic terrestrial toxicity category and the subchronic/chronic human toxicity category. In the subchronic/chronic terrestrial toxicity category, severity factors and safety factors are used as data qualifiers. If studies of duration less than the required duration are used, the NO(A)EL or LO(A)EL is multiplied by a safety factor of 0.3, and the adjusted value is used for scoring. Also, a severity factor is

applied to terrestrial subchronic/chronic toxicity LO(A)EL values to adjust the toxicity values for the severity of the effects observed in the tests. The LO(A)EL (mg/kg/d) is multiplied by a severity factor of 0.1 for severe effects, and by 0.3 for moderate effects. The levels of severity of effects are derived from those described by HARTUNG and DURKIN (1986). Slight effects include: enzyme induction and other reversible biochemical and subcellular changes; reversible hyperplasia, hypertrophy, or atrophy with changes in organ weight; and reversible cloudy swelling and hydropic changes. Moderate effects include: degenerative or necrotic changes with no apparent decrement of organ function, and reversible, slight changes in organ function. Severe effects include: pathologic changes with definite organ dysfunction that are unlikely to be fully reversible, pronounced pathologic changes with severe organ dysfunction with long-term consequences, and death or pronounced life shortening.

Where a LO(A)EL and NO(A)EL from different studies are both available for the same scoring subcategory, the NO(A)EL is preferred unless the adjusted LO(A)EL results in a greater score. Where more than one toxicity value is available for a single subcategory, the value that would result in the greatest score is used to score the subcategory. Information on the subchronic/chronic toxicity of chemicals to rats, mice, guinea pigs, and monkeys is scored under the human health category only since these species often serve as test models for predicting human health effects.

4.2 Basis for criteria and scores

4.2.1 Aquatic life

Criteria and metrics for this scoring system were determined after reviewing existing scoring systems. Most chronic aquatic toxicity scoring systems are similar, with the same number of ranking tiers and nearly the same endpoints. For the five subchronic/chronic aquatic toxicity subcategories (plants, amphibians, cold water fish, warm water fish, and invertebrates) the metrics remained the same, ranging from a maximum of >100 mg/L to a minimum of <0.1 mg/L. The key contributors include the Oak Ridge National Laboratory (O'BRYAN and ROSS, 1988), CMR (MDNR, 1987), OMOE (1990), and the University of Tennessee (DAVIS et al., 1994) systems. The CMR (MDNR, 1987) and Oak Ridge National Laboratory (O'BRYAN and ROSS, 1988) scoring systems had virtually identical chronic aquatic toxicity scoring approaches, with five tiers ranging from >100 mg/L to <0.1 mg/L. The OMOE (1990) and the University of Tennessee (DAVIS et al., 1994) scoring systems gave crucial insight toward the endpoints and criteria required for subchronic/chronic toxicity evaluation.

4.2.2 Terrestrial life

The subchronic/chronic terrestrial toxicity scoring strategy was difficult to develop because there was little agreement among the existing scoring systems regarding metrics, criteria, and acceptable types of data. For mammals, birds, and reptiles and amphibians, a combination of the most restrictive metrics was selected from the CMR (MDNR, 1987), OMOE (1990), Criteria to Identify Candidates for Sunsetting in the Great Lakes Basin (FORAN and GLENN, 1993), and the U.S. EPA Use Clusters (U.S. EPA, 1993) scoring systems. Metrics ranged from >5000 mg/kg/d to <10 mg/kg/d for an available LOEL and from >1000 mg/kg/d to <1 mg/kg/d for an available NOAEL. These scoring systems overlapped to some degree within the five metrics chosen for this scoring system. The CMR's high end ranking and OMOE's low end ranking were used, and Foran and Glenn's method and the U.S. EPA Use Cluster system method were used for the mid-range values. This set of criteria used the greatest number of points from each scoring system to give SCRAM maximum discrimination power in this category. Terrestrial plants and invertebrates were included in the CMR (MDNR, 1987) and OMOE (1990), but the other scoring systems either were unclear about these or did not include them. Since the literature offered no guidance, the CMR (MDNR, 1987) format was applied to these subcategories because its effectiveness had been demonstrated. The CMR (MDNR, 1987) lists the metrics as >100 mg/kg/d to <0.1 mg/kg/d for plants and >5000 mg/kg/d to <10 mg/kg/d for invertebrates. The metrics were adjusted to fit the available data more appropriately, and the units were changed to lbs/acre or kg/ha for terrestrial plants and mg/kg soil for terrestrial invertebrates (see discussion in Section 4.3).

4.3 Specific conditions, concerns, and biases

As in the acute toxicity categories, many difficulties were encountered during the literature search and review process, primarily due to the lack of data. Data availability for subchronic/chronic toxicity, especially for terrestrial plants and invertebrates, was limited since the data did not always fit the scoring requirements. In the case of terrestrial plants and invertebrates, the requirements and endpoints were adjusted to fit the available data. Plant toxicity data were not often available in units of mg/kg soil, and application data were more commonly available. Therefore, the units for terrestrial plant toxicity were changed to kg/ha or lbs/acre. The exposure units for terrestrial invertebrates are mg/kg soil because of the difficulty in finding oral exposure data. Conversion of application data to soil concentrations (and vice versa) was discussed in Section 3.3.

The same difficulties encountered with the terrestrial plant, invertebrate, and reptile and amphibian data under acute toxicity were also experienced when searching for data on subchronic/chronic toxicity for these groups of organisms. More in-depth review should be conducted during the complete risk assessment phase, where individualized and specific judgements may be made about the total weight of evidence of environmental effects observed for a chemical.

5 Subchronic/Chronic Human Toxicity

5.1 Definition and discussion of effects; specific scoring instructions

This category evaluates the potential subchronic/chronic effects on humans. Because of the typical paucity of chemical effects data on humans, this category includes not only epide-

miologic data but also animal bioassay data for rats, mice, guinea pigs, and monkeys, species that are commonly used to evaluate potential toxicologic effects on humans. It is standard practice in biomedical research to use laboratory animals as surrogates for humans in toxicity testing. Although there are disadvantages to using laboratory animal models for prediction of human toxicity, there is no reasonable alternative to the use of animal surrogates. Scoring such animal effects data in this category avoids double counting data from these bioassays as predictive of effects in humans and also as species-specific data under terrestrial subchronic/chronic toxicity.

Studies evaluated in this category consist of repeated dose exposures ranging in duration from several days to years, depending on study protocol and effects to be observed. For instance, a study designed to evaluate developmental effects in rats may include dosing on days 6 through 13 of gestation with an offspring observation period of up to several weeks postpartum. Studies designed to observe longer term effects from low level exposure may run for two years in rats and for seven years or longer in monkeys. Effects observed may range from subtle, reversible enzyme induction to organ system dysfunction resulting in death. Subchronic/chronic toxicity studies used in this subcategory result in the determination of a FEL (frank effect level), LO(A)EL, or NO(A)EL, depending on the effect observed.

Five subcategories have been established for scoring based on the types of effects examined. The "general toxicity" subcategory evaluates typical organ system toxicity not otherwise specifically considered under the other four subcategories. The effects noted typically would include pathology and clinical chemistry observations on hepatic, renal, neurologic, respiratory, cardiovascular, gastrointestinal, and hematopoietic systems. The "reproduction" subcategory focuses on pathology of and physiologic effects on the reproductive system to include not only structure effects but also reproductive viability. The "development" subcategory focuses on embryotoxic, fetotoxic, and teratogenic effects. The remaining two subcategories, "carcinogenicity" and "other toxicity," which are discussed separately, consider neoplastic, mutagenic, immunologic, and potential endocrine effects.

Evaluation of data quality is conducted by comparison of study protocols to recommended study design as defined by the OECD Guidelines for Testing of Chemicals (OECD, 1981) and the U.S. EPA Health Effects Testing Guidelines (U.S. EPA, 1985). Best professional judgement should be used in determining whether the duration of study is sufficient for animals used as surrogates for humans in toxicity testing, bearing in mind the study protocol and effects to be observed, as discussed previously in this section. Repeated dose studies of >28 days, preferably demonstrating a NOAEL, are usually considered to be the minimum requirement. This is consistent with the minimum data requirements for a Tier II value under the Great Lakes Water Quality Guidance (U.S. EPA, 1995b) and with the minimum data requirements for the Organization for Economic Cooperation and Development (OECD) (OECD, 1981) for evaluation of potential human health effects. Because there may be effects that cannot be determined in studies as brief as 28 days, the LO(A)EL or NO(A)EL from a short-term exposure study is multiplied by a safety factor of 0.3 to equate it to an endpoint from a longer-term study. Comparisons of such data, albeit limited, indicate a safety factor of 0.3 should be adequate to adjust to the equivalent of a study of 90 days or greater duration (Weil and McCollister, 1963). Preferences for different types of data and adjustments of toxicity data points for severity are the same as for the subchronic/chronic terrestrial toxicity category.

5.2 Basis for criteria and scores

The criteria and metrics for this category were determined after review of several existing scoring systems. The key contributors include the CMR (MDNR, 1987), OMOE (1990), and Foran and Glenn (1993). The metrics adopted for "effect" levels combine the analogous elements of the CMR (MDNR, 1987) and OMOE (1990) systems. The "no effect" levels are taken directly from the OMOE (1990) system.

Both the CMR (MDNR, 1987) and Foran and Glenn (1993) consider severity of effects in addition to dose in their scoring systems. The CMR (MDNR, 1987) describes effects as severe, moderate, or slight, whereas Foran and Glenn (1993) describe "severity" as high, moderate, or low. The CMR (MDNR, 1987), and apparently Foran and Glenn (1993) as well, follows a severity classification system developed by Hartung and Durkin (1986) and described previously in Section 4.1.1. Although this classification system may best describe general organ toxicity, applications of the concept of slight, moderate, or severe impairment may be comparably applied to reproductive and developmental toxicity, as well as some elements of the "other toxicity" subcategory.

5.3 Specific Conditions, Concerns, and Biases

Subchronic/chronic exposure data for humans are limited, for the most part, to occupational health studies or to limited population studies that are confounded by variable behavior and multiple exposures to other chemical agents, i.e., diet, smoking, alcohol, drugs (prescription, over the counter, nonprescription), workplace and hobby chemical exposures, home cleaning agents, pesticides, cosmetics, etc. It is rare to find sufficiently controlled studies with documented exposure levels adequate enough for hazard scoring. Therefore, reliance on animal bioassay data is necessary. Unfortunately, few chemicals have a well-developed and complete database evaluating carcinogenic, reproductive, developmental and other general toxicologic effects. Therefore, scoring decisions are often made based on data of lesser quality than is desirable.

6 Subchronic/Chronic Human Toxicity – Carcinogenicity

6.1 Definition and discussion of effects; specific scoring instructions

Carcinogenic hazard potential to humans is determined in this scoring system by evaluation of the evidence of increased incidence of tumors, either malignant neoplasms or a combination of malignant and corresponding benign neoplasms,

in humans or laboratory animals exposed to a given chemical. The significance of studies that indicate increased incidence of benign neoplasms without malignant neoplasms is addressed on a case-by-case basis. Knowledge of the mode of action associated with the benign tumor response may assist in determining the significance of benign tumor incidence. Evidence may exist that the observed benign tumors might progress to malignant tumors. It is also possible that benign tumors will have detrimental effects on target tissue function. For example, a benign tumor may damage brain tissue simply because of the space it displaces in the brain case.

Data on humans are available from epidemiologic studies or case reports, and these provide direct evidence that a chemical can cause cancer in humans. Therefore, good quality data on humans are preferred over animal bioassay data for hazard scoring. Unfortunately, there is a relative paucity of data for most chemicals on carcinogenic potential in humans. Further, unless they are based on a high tumor incidence rate, epidemiologic studies usually have a low power to detect carcinogenic response above background. Therefore, if there are no adequate data available for humans, cancer studies in animals may be used to assess hazard potential. Nearly all of the agents known to cause cancer in humans are also carcinogenic in animals (U.S. EPA, 1996).

There are examples in which animals show positive results that are not relevant in humans, e.g., alpha-2u-globulin formation in the male rat kidney. However, in the absence of data to suggest otherwise, positive results in animal studies considering both biological and statistical significance will be used to indicate carcinogenic potential in humans. The technical adequacy of the data from animal studies will be evaluated according to the National Toxicology Program chemical carcinogenesis testing and evaluation principles (National Toxicology Program, 1984) and other similar principles and study guidelines (Office of Science and Technology Policy, 1985; OECD, 1981; U.S. EPA, 1985).

The endpoint used to score carcinogenicity is the ED10 (effect dose 10). To quantify potency, the data are modeled to estimate the dose of a substance associated with a lifetime increased cancer risk of 10%, or the ED10. The potency factor is defined as the reciprocal of the estimated dose, 1/ED10. The ED10 was selected since it is in the observed effects range, and it is relatively insensitive to the choice of dose-response model used. Also, it does not require additional model extrapolation to estimate low dose risk beyond the observed range, and it is a statistically stable estimate (FORAN and GLENN, 1993; U.S. EPA, 1987). Further, because the ED10 is in the observed range, it may be compared more readily with the LO(A)EL or NO(A)EL. The inability to estimate an ED10 necessitates use of an uncertainty score for this subcategory.

For the SCRAM scoring system, the weight of evidence for ascertaining human cancer potential is used to adjust the score based on the ED10, i.e., the greater the likelihood of carcinogenic effects, the greater the score. Weight of the evidence includes not only human epidemiologic data and long term animal bioassay data, but also metabolic and pharmacokinetic properties, structure-activity correlations, effects on the immune and endocrine systems, mutagenic chromosomal or DNA interaction effects, etc. Such detailed review of all data in weight of evidence considerations for carcinogenic hazard assessment is generally outside of the design of this simple scoring system and best considered during a detailed chemical risk assessment. However, major observations such as biological appropriateness of effects to humans, structural similarity to known carcinogens, strong mutagenic evidence, etc., are necessary to determine the strength of the evidence and are used in this system.

SCRAM uses the weight of evidence descriptors from the Proposed Guidelines for Carcinogen Risk Assessment (U.S. EPA, 1996), which suggest three major categories for describing human cancer potential: "known/likely," "cannot be determined," and "not likely." There are several subcategories within the three main categories. "Known" to be carcinogenic in humans is based on epidemiologic evidence or a combination of epidemiologic and experimental evidence demonstrating causality between human exposure and cancer. "As if known" to be carcinogenic in humans is based on a combination of epidemiologic data showing a plausible association and strong experimental animal evidence. "Likely" to produce cancer in humans is due to evidence of tumor production in animals by modes of action relevant to human carcinogenicity. "Cannot be determined" is the category for tumor effects or other key data that are suggestive, conflicting, or limited in quantity and therefore inadequate to convincingly demonstrate human cancer potential. Subcategories for "cannot be determined" currently include (1) "suggestive evidence" that raises concern for carcinogenic effects, (2) "conflicting data" in which some evidence suggests carcinogenic effects but other pertinent evidence does not confirm concern, (3) "inadequate," and (4) "no data." "Not likely" describes the level of experimental evidence that is satisfactory for deciding that there is no basis for cancer concern.

The ED10 is evaluated against the dosage scale of the system and a score determined. Based on the weight of evidence review, a multiplier is applied to the ED10 dosage level to enhance the scores of those chemicals demonstrating a greater weight of evidence towards human carcinogenicity. This metric was established to move the score up one scoring unit for "known" human carcinogens and up two-thirds of a unit for "likely" human carcinogens. No adjustment is proposed for "suggestive" or "conflicting" evidence. The higher score reflects an increased concern for human carcinogenicity as a severe and dreaded effect and also reflects a similar adjustment for "severe" effects used in the other subchronic/chronic toxicity categories.

6.2 Basis for criteria and scores

The criteria and metrics for this scoring system were developed after reviewing several existing scoring systems. The key contributors include the CMR (MDNR, 1987); Foran and Glenn (1993); CERCLA Section IIIC. Summary of the Methodology for Adjusting the RQs of Potential Carcinogens (U.S. EPA, 1987); CERCLA Section IIB. Reportable Quantity Adjustment Methodology (U.S. EPA, 1989a); the Use Clusters Scoring System (U.S. EPA, 1993); and review

of the recent Proposed Guidelines for Carcinogen Risk Assessment (U.S. EPA, 1996).

CERCLA (USEPA, 1987, 1989a), the Use Clusters system (U.S. EPA, 1993), and Foran and Glenn (1993) incorporate cancer potency with the weight of evidence. The weight of evidence descriptors from the Proposed Cancer Guidelines (U.S. EPA, 1996) used in this scoring system replace the old U.S. EPA classifications of "known," "probable," and "possible" human carcinogens, or A, B, and C level carcinogens. This scoring system divides the weight of evidence in a manner similar to that used in the old system. "Known" and "as if known" are considered comparable to Class A carcinogens (human carcinogens). "Likely" is considered comparable to Class B carcinogens (probable human carcinogens) as well as to a select group of Class C carcinogens (possible human carcinogens). "Cannot be determined," "suggestive evidence," and "conflicting data" subdivisions generally correspond to Class C carcinogens (possible human carcinogens).

The metrics selected for scoring are a modification of the metrics in the scoring systems developed by Foran and Glenn (1993) and in the U.S. EPA Use Clusters system (U.S. EPA, 1993). Metrics were modified to encompass five dosage levels to match the rest of the scoring systems. The break points selected for the various scores reflect an assessment of ED10 values for numerous carcinogens on the list of Great Lakes contaminants. An attempt was made to develop a graded distinction among the listed carcinogens, i.e., to spread the scores out more evenly over a wider range than is used by Foran and Glenn (1993) and the Use Clusters (U.S. EPA, 1993) process. The dosages selected are arbitrary to meet that objective.

6.3 Specific conditions, concerns, and biases

Carcinogenicity scoring depends totally on the availability of human epidemiologic or whole animal bioassay data. Short-term assays predictive of cancer potential are used only as supportive information in developing a weight of evidence assessment. Mutagenicity data are specifically discussed in the "other toxicity" section. ED10 values are generally identified from a linearized multistage model (LMS) assessment of animal bioassay data.

7 Subchronic/Chronic Human Toxicity – Other Toxicity

7.1 Definition and discussion of effects; specific scoring instructions

The "other toxicity" subcategory is designed to address evidence of potential adverse effects on humans that cannot be addressed adequately by the other subcategories. Such evidence may include *in vitro* assay data on mutagenicity, estrogen-like influences that may impact the endocrine system, behavioral effects, immune system effects, etc. These effects have less standardized test protocols, are areas of developing science, and often lack clear correlations to whole animal effects and dose-response relationships that can be translated to whole animal exposures. As a result, correlation to and inference of effects on humans tends to be qualitative rather than quantitative and based on strength of evidence rather than clear dose-response cause-and-effect relationships. For these reasons, this subcategory acts more as a modifier for the potential lack of adequate whole animal or epidemiologic evidence that fits the other subcategories; i.e., there may be data here that can be used *in lieu* of an uncertainty score for the category. It also provides an opportunity to demonstrate other potential adverse effects to humans that would not be identified as a result of the standardized methodologies/protocols of the other subcategories.

A major effect to be considered in this subcategory is mutagenicity. The greatest importance of this subcategory is to identify chemicals which may adversely affect the germinal tissue causing germ-line mutations that may be passed on to future generations. This can be determined either directly with an *in vitro* germ-cell mutation test or with positive evidence of mutagenic potential and chemical interaction with mammalian gonad tissue.

Lesser weights of evidence categories towards germ-line mutagenicity are, in descending order: "possible germ-line mutagen" with some evidence of germ-line mutagenicity, "positive somatic-cell mutagen" with positive evidence of somatic cell mutagenicity, and "possible somatic-cell mutagen" with suggestive evidence of somatic cell mutagenicity. A more complete discussion of the definitions and scoring criteria for this subcategory is given in the Michigan Critical Materials Register Criteria and Technical Support Document (MDNR, 1987).

Immune system effects for this subcategory may include immunosuppression, alterations of host defense mechanism against pathogens or neoplasia, and allergies or autoimmunity. A fourth effect, uncontrolled proliferation such as in leukemia and lymphoma, is considered under carcinogenicity.

Behavioral alterations under "other toxicity" may include changes in sensations of sight, hearing, or touch; changes in simple or complex reflexes and motor functions; changes in cognitive functions such as learning, memory, or attention; changes in mood such as fear or rage; disorientation as to person, time, or place; or distortions of thinking and feeling such as delusions and hallucinations. Unfortunately, relatively few neurotoxic syndromes in terms of initial neurochemical change, structural alterations, physiological consequence, and behavioral effects have been thoroughly characterized (U.S. EPA, 1995a).

Human behavioral toxicology has adopted the neurologic/neuropsychologic model. Information from neurological exams is mostly qualitative and descriptive rather than quantitative, limiting its usefulness for neurotoxicologic risk assessment. Estimates of severity of functional impairment can be placed reliably in only three or four categories (i.e., mild, moderate, and severe). Much depends on the subjective judgement of the examiner. The neuropsychological battery test also requires the interpretation of a trained neuropsychologist. It is not very sensitive to low-level effects and relies on the individual or family to identify signs and symptoms (U.S. EPA, 1995a).

As compared with human studies, animal studies are more often available, provide more precise exposure, and have

better control over environmental factors. Therefore, emphasis on animal data may provide more appropriate results for estimating potential for neurobehavioral effects. Endpoints for behavioral effects testing according to Toxic Substances Control Act (TSCA) (U.S. EPA, 1985) and Federal Insecticide, Fungicide and Rodenticide Act (FIFRA) (U.S. EPA, 1991) test guidelines include the use of functional observational batteries (FOB), motor activity, and schedule-controlled behavior. In addition, there are many other measures of behavior and other test methods that may be considered. FOB are designed to detect and quantify major overt behavioral, physiological, and neurological signs. Some of these tests are similar to chemical neurological observations used with humans and therefore are subject to the same shortcomings as the human assessment methods. Discussions of these test methods and endpoints are detailed in U.S. EPA's Proposed Guidelines for Neurotoxicity Risk Assessment (U.S. EPA, 1995a). Identification of a critical adverse effect often requires considerable professional judgement.

Due to concern for possible adverse human and wildlife impacts from anthropogenic chemicals causing interference with normal endocrine functions, increased research on endocrine effects has occurred in recent years. The "endocrine disrupter" hypothesis is that exogenous (externally produced) substances can sometimes mimic endogenous (internally produced) hormones, activating similar responses, blocking or diminishing the function of endogenous hormones by competing for available receptor binding sites, or acting directly without receptor mediation. This is a very complex process to evaluate given that humans and wildlife are exposed to complex mixtures of natural exogenous hormones, anthropogenic hormonal mimics and endocrine modulating conditions with different effects and potencies. Depending on dose and combination, these chemicals may elicit a wide variety of effects. Further, many of the test protocols developed to study this issue do not follow the typical whole animal bioassay structure/function endpoint assessment protocols.

Many of the endocrine disruption test protocols are *in vitro* assays that estimate the estrogenic activity of a chemical relative to 17β-estradiol. Some examples are the human breast tumor MCF-7 cell proliferation assay (SOTO et al., 1991), the rainbow trout hepatocyte culture assay for vitellogenin secretion (WHITE et al., 1994), and the estradiol receptor binding assay using rainbow trout liver cytosolic extract (WHITE et al., 1994). The significance of the results of these assays relative to whole animal effects has yet to be determined. The results of these studies are reported here in the "other toxicity" subcategory as additional information to be considered in evaluating the potential for endocrine system effects.

A score is calculated for each of the areas in this subcategory. Since this subcategory is to be informational as a modifier to increase toxicity concern for the chemical, no uncertainty point is assigned when data are lacking. The greatest of the scores for mutagenicity, behavioral effects, immune system effects, and endocrine effects is the score assigned to the subcategory. No severity modifier is used to adjust the score in this subcategory since, by definition, severity is taken into account in the criteria.

7.2 Basis for criteria and scores

The criteria for this subcategory are more qualitative than those for the other scoring categories and require narrative descriptions that relate to weight of evidence or severity of effects. For mutagenicity, the key contributor to the method selected for SCRAM was the CMR (MDNR, 1987). The criteria from the CMR (MDNR, 1987) have been adopted directly, with the scores for the various criteria elements adjusted to fit SCRAM. For the other effects, scoring becomes more subjective and dependent on professional judgement as to severity. Critical factors to be considered include immunotoxicity, behavioral effects, and the potential for adverse effects to occur in whole animal systems based on evidence from *in vitro* assays, e.g., endocrine mimicking effects. For immunotoxicity and behavioral effects observed at reasonable exposure levels (1% of daily diet), scores are to be assigned from maximum score to minimum score in the following order: severe, irreversible effects; severe, yet potentially reversible or moderate, irreversible effects; moderate, reversible effects; and slight/no effects. For potential endocrine system effects, given the current lack of direct correlation between *in vitro* assay findings and comparable whole animal effects, the greatest score that will be assigned is 3, meaning that the chemical has a "strong potential" to cause disruptive endocrine system effects. The next level is "moderate potential," and the least level is "weak (minimal) potential." Qualitative scores for immunotoxicity and behavioral effects will be defined according to endpoints defined by OECD (1981), TSCA (U.S. EPA, 1985), FIFRA (U.S. EPA, 1991) test protocols, and U.S. EPA Proposed Guidelines for Neurotoxicity Risk Assessment (U.S. EPA, 1995a). The potential estrogenic effects of chemicals are determined relative to the potency of 17β-estradiol. Other parameters used to identify endocrine modulation effects also must be measured against a standard to determine the biological significance of these *in vitro* assay effects to whole animal toxicity.

7.3 Specific conditions, concerns, and biases

For this subcategory, the quality and quantity of data available are variable. This variability is the result of many years of testing without standardized protocols. Best professional judgement will be necessary in determining the suitability of data for scoring. Study protocols will be compared, as much as is feasible, to currently acceptable test methods. For example, dose-response data from multiple dose studies and results with greater statistical significance will receive additional weight in scoring. However, given the lack of standardization of some study protocols, it may be necessary to use a weight of evidence approach to determine the significance of these study findings to whole animal effects.

Acknowledgments

This project was completed through the combined efforts of the Surface Water Quality Division, Michigan Department of Environmental Quality and the Department of Fisheries and Wildlife, Michigan State University. Assistance and advice in developing the chemical ranking and scoring system was provided by an advisory committee of toxi-

cologists, chemists, and biologists from several Michigan industrial organizations, environmental groups, and academicians, as well as technical representatives from U.S. EPA Region V, the Ontario Ministry of Environment and Energy, the State of Wisconsin, and the State of Minnesota. Funding for this project was provided through a grant from the Michigan Great Lakes Protection Fund, administered by the Office of the Great Lakes, Michigan Department of Environmental Quality.

8 References

ASTM (1992a): Standard guide for conducting acute toxicity tests with fishes, macroinvertebrates, and amphibians. Designation E 729-88a. Pages 403-422. In: 1992 Annual book of ASTM standards. Section 11. Water and environmental technology. Volume 11.04. Pesticides; resource recovery; hazardous substances and oil spill responses; waste management; biological effects. Publication code # 01-110492-48. American Society for Testing and Materials, Philadelphia, PA

ASTM (1992b): Standard guide for conducting early life-stage toxicity tests with fishes. Designation E 1241-92. Pages 886-913. In: 1992 Annual book of ASTM standards. Section 11. Water and environmental technology. Volume 11.04. Pesticides; resource recovery; hazardous substances and oil spill responses; waste management; biological effects. Publication code # 01-110492-48. American Society for Testing and Materials, Philadelphia, PA

ASTM (1992c): Standard guide for conducting renewal life-cycle toxicity tests with *Daphnia magna*. Designation E 1193-87. Pages 866-882. In: 1992 Annual book of ASTM standards. Section 11. Water and environmental technology. Volume 11.04. Pesticides; resource recovery; hazardous substances and oil spill responses; waste management; biological effects. Publication code # 01-110492-48. American Society for Testing and Materials, Philadelphia, PA

Foran, J.A.; Glenn, B.S. (1993): Criteria to identify chemical candidates for sunsetting in the Great Lakes Basin. The George Washington University, Environmental Health and Policy Program, Department of Health Care Sciences, Washington, DC

Hartung, R.; Durkin, P.R. (1986): Ranking the severity of toxic effects: Potential applications to risk assessment. Comments Toxicol. **1**, 49-63

MDNR (1987): Michigan Critical Materials Register criteria and support document. Michigan Department of Natural Resources, Surface Water Quality Division, Lansing, MI

OMOE (1990): Ontario MOE scoring system. Ontario Ministry of the Environment, Hazardous Contaminants Coordination Branch. Toronto, Ontario

National Toxicology Program (1984): Report of the Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation of National Toxicology Program, Board of Scientific Counselors. U.S. Government Printing Office, Washington, DC, 1984-421-132:4726

O'Bryan, T.R.; Ross, R.H. (1988): Oak Ridge National Laboratory. 1988. Chemical scoring system for hazard and exposure identification. J. Toxicol. Environ. Health **1**, 119-134

OECD (1981): Guidelines for testing of chemicals. Organization for Economic Cooperation and Development. Paris, France

Office of Science and Technology Policy (1985): Chemical carcinogens: Review of the science and its associated principles. Federal Register **50** (March 14, 1985), 10372-10442

Rand, G.M. (Ed.) (1995): Fundamentals of aquatic toxicology: Effects, environmental fate, and risk assessment. 2nd edition. Taylor & Francis, Washington, DC, pp. 1125

Snyder, E.M.; Snyder, S.A.; Giesy, J.P.; Blonde, S.A.; Hurlburt, G.K.; Summer, C.L.; Mitchell, R.R.; Bush, D.M. (1999a): SCRAM: A scoring and ranking system for persistent, bioaccumulative, and toxic substances for the North American Great Lakes. Part I. Structure of the scoring and ranking system. ESPR – Environ. Sci. & Pollut. Res. Online Publication (Online First): October 29th, 1999. Print Publication: ESPR **7** (1) 2000

Snyder, E.M.; Snyder, S.A.; Giesy, J.P.; Blonde, S.A.; Hurlburt, G.K.; Summer, C.L.; Mitchell, R.R.; Bush, D.M. (1999b): SCRAM: A scoring and ranking system for persistent, bioaccumulative, and toxic substances for the North American Great Lakes. Part II. Bioaccumulation potential and persistence. Acute and subchronic or chronic toxicity. ESPR – Environ. Sci. & Pollut. Res. Online Publication (Online First): October 29th, 1999. Print Publication: ESPR **7** (3) 2000

Snyder, E.M.; Snyder, S.A.; Giesy, J.P.; Blonde, S.A.; Hurlburt, G.K.; Summer, C.L.; Mitchell, R.R.; Bush, D.M. (1999c): SCRAM: A scoring and ranking system for persistent, bioaccumulative, and toxic substances for the North American Great Lakes. Part IV. Results from representative chemicals, sensitivity analysis, and discriminatory power. ESPR – Environ. Sci. & Pollut. Res. Online Publication (Online First): October 29th, 1999. Print Publication: ESPR **7** (4) 2000

Soto, A.M.; Justica, H.; Wray, J.W.; Sonnenschein, C. (1991): p-Nonylphenol: an estrogenic xenobiotic released from "modified" polystyrene. Env. Health Perspect. **92**, 167-173

U.S. EPA (1985): Toxic Substances Control Act test guidelines; final rules. Federal Register **50**(188) (September 27, 1985), 39421-39425

U.S. EPA (1987): Reportable Quantity adjustments; proposed rule. Federal Register **52**(50) (March 16,1987), 8140-8186

U.S. EPA (1989a): Reportable Quantity adjustments; final rule. Federal Register **54**(155) (August 14,1989), 33418-33484

U.S. EPA (1989b): Technical background document to support rule making pursuant to CERCLA Section 102. Volume 3. United States Environmental Protection Agency, Office of Solid Waste and Emergency Response. Washington, DC

U.S. EPA (1991): Pesticide assessment guidelines, subdivision F. Hazard evaluation: Human and domestic animals. United States Environmental Protection Agency, Office of Pollution Prevention and Toxics. Washington, DC

U.S. EPA (1993): Chemical use clusters scoring methodology. United States Environmental Protection Agency, Office of Pollution Prevention and Toxics, Chemical Engineering Branch, Washington, DC. pp. 57

U.S. EPA (1995a): Proposed guidelines for neurotoxicity risk assessment; public notice. Federal Register **60**(192) (October 4, 1995), 52032-52056

U.S. EPA (1995b): Water quality guidance for the Great Lakes system; final rules. Federal Register **60**(56) (March 23, 1995), 15366-15425

U.S. EPA (1996): Proposed guidelines for carcinogen risk assessment. United States Environmental Protection Agency. Federal Register **51**(185), September

Weil, C.S.; McCollister, D.D. (1963): Relationship between short and long term feeding studies in designing an effective toxicity test. Agr. Food Chem. **11**(6), 486-491

White, R.; Jobling, S.; Hoare, S.A.; Sumpter, J.P.; Parker, M.G. (1994): Environmentally persistent alkylphenolic compounds are oestrogenic. Endo. **135**, 175-182

Received: February 19th, 1999
Accepted: September 6th, 1999
Online Publication (Online First): October 29th, 1999